International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Classification of Mango Fruit Varieties using Naive Bayes Algorithm

Ohnmar Win

Lecturer, Department of Electronic Engineering, Mandalay Technological University, Mandalay, Myanmar

How to cite this paper: Ohnmar Win "Classification of Mango Fruit Varieties using Naive Bayes Algorithm" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1475-1478, **https://doi.org/10.31142/ijtsrd26677**

IJTSRD26677

ABSTRACT

Mangos are an important agricultural commodity in the global market for fresh products. In Myanmar, the type of mango called SeinTaLone is the best taste and the most people like it. Another type of mango called MaSawYin is not good taste but it is visually similar to the SeinTaLone. So, some people are difficult to classify the mango varieties. A means for distinguishing mango varieties is needed and therefore, some reliable technique is needed to discriminate varieties rapidly and non-destructively. The main objective of this research was to classify the varieties of mango fruit that occur in Myanmar using Naive Bayes algorithm. The methodology involved image acquisition, pre-processing and segmentation, feature extraction and classification of mango varieties. A method for classifying varieties of mangos using image processing technique is proposed in this paper. RGB image was first converted to HSV image. Then by using edge detection method and morphological operation, region of interest was segmented by taking into account only the HUE component image of the HSV image. Later, a total of 4 shape features and 13 texture features were extracted. Extracted features were given as inputs to a Naive Byaesian classifier to classify the test images as each type. The data set used had 50 mango images for each varieties of mango for training and 20 images of mango for each variety for testing.

KEYWORDS: *HSV, Edge Detection, Features Extraction, Naive Bayesian classifier*

I. INTRODUCTION

Mangos are one of the most commonly consumed fruits in the world. The quality of a mango depends on its external characteristics, such as color, size, and surface texture, and internal parameters, such as sweetness, acidity, firmness, tissue texture, ascorbic acid, and polyphenolic compounds. These characteristics, especially internal and external parameters, are similar to a variety. However, each variety has its special characteristics and flavor, which results in different prices and preferences by different people.

Mango produce dealers have warehouses that store different varieties of mango fruits. Therefore, different mango varieties easily get mixed up during harvesting, storage and marketing. Most mango produce dealers will sort the mangos manually which results in high cost, subjectivity, tediousness and inconsistency associated with manual sorting. The main objective of this study was to investigate the applicability and performance of Naive Bayes algorithm in the classification of mango fruit varieties.

The Automated Fruit Classification System is embedded as well as image processing based totally automated system. This system is very useful to the farmers. Fruit classification system is a totally automated and due to that it saves the valuable time of the farmer as well as the buyers and customers. This system reduces the labor intensity and increases the quality of the fruit.

II. Methodology

This system is implemented for Mango fruit varieties classification system with image processing techniques. Implementation of the system is worked out with the help of MATLAB and Camera. The main techniques are color conversion, edge detection for image segmentation, features extraction and Naive Bayes classifier. Figure 1 shows the system flow diagram. Following are the steps in the approach:

- Capturing the fruit images
- Preprocessing
- Conversion of each image RGB to HSV color image.
- Segment the fruit region using edge detection
- Features extraction
- Classification

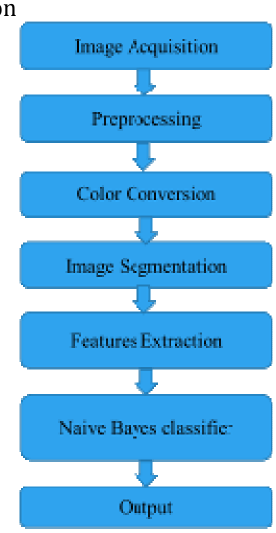

Figure1. System Flow Diagram

A. Image Acquisition
The proposed system can classify five verities of mango fruits such as SeinTaLone, MaSawYin, HinThar, PanSwae and YinKwal. The mango verities images are captured with JPEG format by using a phone camera and 1-3 feet distance. And, fruit images are taken from the white background or plain background at day time or night. These images were cropped into smaller images and stored in JPEG format. The acquired mango varieties images are shown in Fig 2. The experimented mango varieties that are occur in Myanmar included: PanSwae, HinThar, MaSawYin, YinKwal and SeinTaLone.

(a) (b) (c) (d) (e)

Figure2. Mango varieties: (a) PanSwal (b) HinThar (c) MaSawYin (d) YinKwal (e) SeinTaLone

B. Preprocessing
In order to get mango features accurately, mango fruits images were pre-processed through different pre-processing methods. These methods were resized the images and filtering the images to remove noise as described below.

1. Resizing the image
After loading the input image, it is resized into (300,400). In order to prevent distortion of the image, the smallest dimension of the image is expanded with zero-value rows or columns of pixels on both sizes.

2. Filtering
The median filter was implemented in this process to remove noise. The median filter computes the median value of the gray-scale values within a rectangular filter window surrounding each pixel. This has the effect of smoothing the image (eliminating noise).

C. Segmentation
The segmentation and pre-processing task are the initial stages before the image is used for the next process. The main objective of this process is to obtain the binary image.

1. Conversion of each image RGB to HSV color image
Fruit images are usually captured in RGB color space, however, many works in the literature discovered that other color spaces such as HSV, Lab could be more useful than RGB in the extraction of fruit region. Therefore, RGB image is converted to HSV image in this system. HSV color space highlights the fruit of interest and makes it more prominent then other components. It makes more efficient the fruit localization process and then more intuitive about color of brightness and spectral name than the mixture coefficients of RGB.

$$\begin{cases} 0 & if\ c=0 \\ 60^\circ * \left(\frac{g-b}{c}+0\right) & if\ c=r \\ 60^\circ * \left(\frac{r-b}{c}+2\right) & if\ c=g \\ 60^\circ * \left(\frac{r-g}{c}+4\right) & if\ c=b \end{cases} \quad (1)$$

$$V = M \quad (2)$$

$$S = \begin{cases} 0 & if\ c=0 \\ \frac{c}{M} & if\ c \neq 0 \end{cases} \quad (3)$$

Where, r= normalized value of red, g=normalized value of green, b=normalized value of blue, M= maximum value of RGB, m= minimum value of RGB.

2. Edge detection
Edge detection is a process that detects the presence and location of edges constituted by sharp changes in intensity of an image. Edges define the boundaries between regions in an image, which helps with segmentation and object recognition. Mango fruit images are changed to binary (black and white) format since edge detection can be done on binary (black and white) or grey scale images. The binary images are detected the edges by using Sobel mask operator.

The Sobel edge detection operation extracts all of edges in an image, regardless of direction. Sobel operation has the advantage of providing both a differencing and smoothing effect. It is implemented as the sum of two directional edge enhancement operations. The resulting image appears as an unidirectional outline of the objects in the original image. Constant brightness regions become black, while changing brightness regions become highlighted. Derivative may be implemented in digital form in several ways. However, the *Sobel operators* have the advantage of providing both a differencing and a smoothing effect. Because derivatives enhance noise, the smoothing effect is particularly attractive feature of the Sobel operator [1] .

D. Feature Extraction
In the presented method, texture and shape features were extracted from the mango images which were used as inputs for classification by being fed into Naive Bayes algorithm. Brief information is provided for these features as follow.

1. Geometric or Shape Features
According to hematologists, the geometric of the fruit is one of the essential features which can be used for classification of the fruits. Geometric features provide information about the size and shape of a fruits. They are computed from the fruits binary image. We used 4 geometric features including:

- Area – the total number of non-zeros pixels available within the image region.
- Perimeter – the distance between successive boundary pixels.
- Major Axis length– The length of the line which connects the two farthest
- Minor Axis – The length of the line connecting the two closest boundary points

2. Statistical or Texture Features
In this research, two different statistical-based methods are selected for texture feature extraction, namely Histogram-based approach and Gray level Co-occurrence Matrix (GLCM). The histogram {h}of an image is calculated based on the frequency occurrence of each individual gray-level intensity value in the image [3][4].

The texture features based on the image histogram can be computed as follow.

$$Mean, \mu = \sum_{i=0}^{N-1} i\, h(i) \quad (4)$$

$$Standard\ deviation, \sigma = \sum_{i=0}^{N-1} \left(i\ \mu(h)\right)^2 h(i) \quad (5)$$

$$Skewness, \mu_3 = \frac{\sum_{i=0}^{N-1}(i-\mu)^3 h(i)}{\sigma^3} \qquad (6)$$

$$Kurtosis, \mu_4 = \frac{\sum_{i=0}^{N-1}(i-\mu)^4 h(i)}{\sigma^4} - 3 \qquad (7)$$

Gray Level Co-occurrence Matrix (GLCM) [2] is one of the most powerful and popular statistical texture analysis methods for extracting texture information from an image. Texture features based on the GLCM can be computed as follow.

Homogeneity – the closeness of the distribution of elements to the diagonal.

$$\text{Homogeneity} = \sum_i \sum_j \frac{1}{1+|i-j|} * p(i,j) \qquad (8)$$

Energy – to measure uniformity of the normalized matrix.

$$\text{Energy} = \sum_{i=0}^{N-1} h^2(i) \qquad (9)$$

Correlation – correlation between pixel values and its neighborhood.

$$\text{Correlation} = \sum_i \sum_j \frac{(i-u_x)(j-u_y)\,p(i,j)}{\delta_x \delta_y} \qquad (10)$$

Entropy – to measure the randomness of intensity distribution.

$$\text{Entropy} = \sum_i \sum_j p(i,j)\log(p(i,j)) \qquad (11)$$

Contrast – type of opposition between two objects, highlighted to emphasize their differences.

$$\text{Contrast} = \sum_i \sum_j |i-j|^2 * p(i,j) \qquad (12)$$

E. Naive Bayes

Naive Bayes classifier is a probabilistic classifier based on the Bayes theorem, considering Naive (Strong) independence assumption. Naive Bayes classifiers assume that the effect of a variable value on a given class is independent of the values of another variable. This assumption is called class conditional independence. Naïve Bayes can often perform more sophisticated classification methods. It is particularly suited when the dimensionality of the inputs is high. When we want more competent output, as compared to other methods output we can use Naïve Bayes implementation. Naive Bayes is used to create models with predictive capabilities.

$$Bayes'\,Theorem: Probability\,(B\,given A) = Probability\,\frac{A\ and\ B}{Probability}$$

III. Test and Results

The proposed system has two main stages: Training and Testing. In the training stage , there was an mango varieties image database consisting of 70 samples of each type of mango which were used for training, validation, and testing purposes. After training the system, it then classified the mango varieties whether it is PanSwae, HinThar, MaSawYin, YinKwal and SeinTaLone during testing and validation stages. The step by step testing results are shown in the follow.

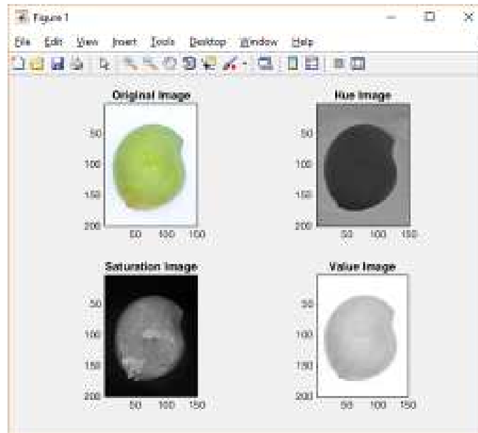

Figure3. Mango Image Color Transformation

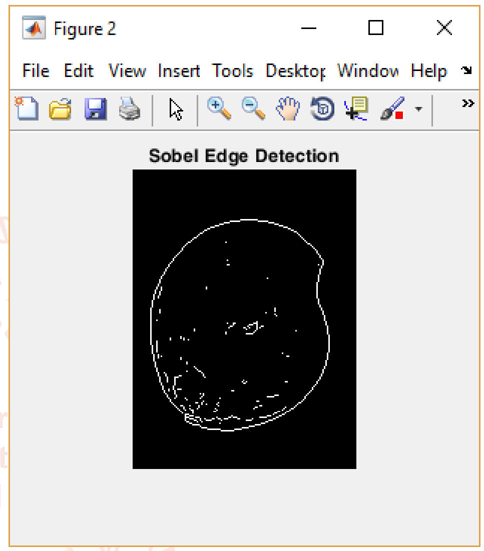

Figure4. Edge Detection

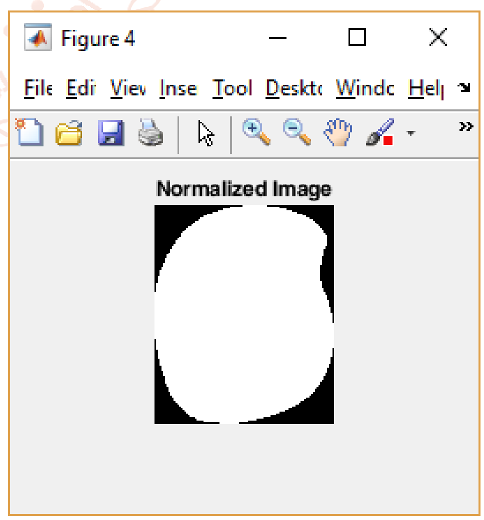

Figure5. Segmented Image

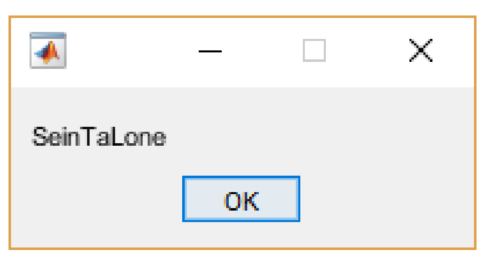

Figure6. Classification Results

For the best classification system, the accuracy should be high. It is calculated as ratio of number of correctly recognised fruit samples upon total number of samples used in testing. The performance for the system is evaluated using five varities of fruit whose samples are not in dataset. The samples are tested with our system. The proposed system is implemented for testing 20 images for each mango variety. The 50 images for each variety of mango fruit are used to train in the database.

$$Accuracy = \frac{\text{No.of Correctly classified Frut Samples}}{\text{Total No.of Fruit Samples}} \times 100 \quad (13)$$

Table1. Result for Testing Data Set

Sr. No.	Fruit Name	Total Samples	No. of Fruit correctly classified	Not Recognized	Accuracy (%)
1	PanSwae	20	19	1	95
2	HinThar	20	20	0	100
3	MaSawYin	20	18	2	90
4	YinKwal	20	19	1	95
5	SeinTaLone	20	18	2	90

The accuracy of the system is calculated with the help of equation 13 and the correctly classified samples and the samples which are not correctly classified are shown by values in the table 1.

IV. Conclusion

The accuracy for the testing data set showed that the highest accuracy in Naive Bayes was observed in HinThar (100%). The lowest accuracy is MaSawYin and SeinTaLone (90%). YinKwal and PanSwae has (95%) accuracy rate. HinThar had a specificity of 0% because during validation and testing we did not find its true negative and false positive values. This can be attributed to its shape which was easily distinguishable from the other mango varieties whose shapes were almost similar. The average accuracy for the system was 94% for the testing data set. The study indicated that Naive Bayes has good potential for identifying apple varieties nondestructively and accurately.

References

[1] Srikanth Rangarajan, "Algorithms For Edge Detection",

[2] Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics, SMC-3*(6),610-621. doi: 10.1109/tsmc.1973.4309314

[3] Selvarajah, S., & Kodituwakku, S. (2011). Analysis and comparison of texture features for content based image retrieval. *International Journal of Latest Trends in Computing,2*(1).

[4] Srinivasan, G., & Shobha, G. (2008). Statistical Texture Analysis. *Proceedings of World Academy of Science: Engineering & Technology, 48*.